Focus 2 Research

PROJECT PROPOSAL on DATA ARCHIVES/REPOSITORIES

RESEARCH RATIONALE:
The bibliography and abstracting activities of the IP2 Scientific Focus (Focus 2) would benefit by the addition of information about actual practices, standards, and protocols currently in place at existing data archives / repositories / catalogues / portals in the sciences. The access and dissemination of data in the sciences relies heavily on these types initiatives that may or may not include archiving or preservation as a mandate.

PURPOSE:
The purpose of this project is to compile structured information about the standards and protocols in place at data archives / repositories / catalogues / portals in the sciences. The information will serve Focus 2, the policy cross-domain, and the description cross-domain.

METHOD:
The information will be acquired in a brief survey, to be undertaken by a UBC research Assistant according to the general directions outlined below. The survey will be preliminary in nature and is not intended to be comprehensive and exhaustive.

OVERALL GOALS:
To provide concrete information about the understanding of issues of accuracy/reliability/authenticity in the sciences

To collect information about the ways and means that are used to underwrite confidence in the accuracy/reliability/authenticity as practiced and implemented in the sciences (natural, physical, engineering and etc. and not social sciences).

To understand the structures in place at the surveyed data archives / repositories / catalogues / portals in the sciences that data providers and users rely upon to ensure access, accuracy accuracy/reliability/authenticity of the data.

To provide the IP2 Policy Cross-Domain with examples of procedures and practices from communities of practice in the sciences.

To understand community of practice in scientific data regimes.

To provide the IP2 Description Cross-Domain with examples of metadata standards and warrant from communities of practice in the sciences.

To understand how assemblages of data in the sciences, in data portals / archives / catalogues / directories / centers, address and demonstrate that a proven lineage and a presumption of reliability and authenticity are built in to their process of data acquisitions / data ingestion processes.

To develop and understanding via a small project about the types of holdings and warrants for those holdings in the sciences

DESCRIPTION OF THE PROJECT:
The project will assemble a list of data archives and/or repositories / access portals / catalogues / data centres in the sciences (exclusive of the social sciences) and indicate the following information about their holdings and processes:

1. Name of data science archive / repository / portal / catalogue / centre:

2. Contact information (address, country, URL, etc.):

3. Date the data science archive / repository / portal / catalogue / centre was established:

4. Span dates data science archive / repository / portal / catalogue / centre holdings:

5. Mandate or vision statement of the data science archive / repository /portal / catalogue / centre:

6. Does this data science archive / repository / portal / catalogue / centre have partners and collaborators? If so please list:

7. The community this data science archive / repository / portal / catalogue / centre aims to serve (community of practice, discipline, Gov., etc.)

8. Area or region of interest of the data science archive / repository /portal / catalogue / centre (Antarctica, Africa, Ontario, City, Ecoregion, Watershed, etc.).

9. The type of institution that manages the data science archive / repository/ portal / catalogue / centre (University research centre, scientific association, government, NGO, private sector and etc.).

10. The data science archive / repository / portal / catalogue / centre source of funding?

11. A general description of the data that is held and the sources for the Data (geographic, topographic, health, etc.)?

12. Where are the archive / portal / repository /catalogue / centre data stored (e.g., the scientists keeps the data & registers it, the agency both catalogues and stores most of the data, etc.)

13. If not an archive, does the portal / repository /catalogue / centre have a statement related to archiving?

14. # of data sets?

15. Are data formats specified?

16. How is data accessed? (e.g., users have to register, membership, catalogue, search a list of links, can download, can order CDs, PDF document, order forms, etc.)

17. Is there an access fee service or a data purchase fee?

18. Any statements that indicates how the data science archive / repository /portal / catalogue / centre ensures the quality, accuracy or reliability of its data (ex. Fields found in metadata)

19. What metadata standards if any are used to register or catalogue data in the data science archive / repository / portal / catalogue / centre?

20. What is the process by which a data provider includes his or her organization's data into the data science archive / repository / portal / catalogue / centre? (ex. Quality assurance)

21. Is there an authenticity statement? (ex. Lineage & institution)

22. Are there any terms and conditions of use? Costs? and/or disclaimers?

DESCRIPTION OF THE JOB:
The student researcher should:

1. Develop a search strategy and check with the supervisors before proceeding. (4 hours)

2. Develop a format for reporting results. (3 hours)

3. Following confirmation, undertake a search as a pilot and report results. (3-4 hours)

4. Following any changes to the strategy and/or format, undertake the search and assemble the data. (21 hours)

5. Report results and after feedback, complete project (3 hours)= 35 hours or so)

PROGRESS to DATE:

- Stephen Gage, UBC IP2 Student was allocated to this activity.
- A survey template was developed
- 13 entries have been completed by Stephen Gage
- Sherry Xie, is currently allocated to this activity
- Focus 2 members have submitted a list of data archives / repositories / catalogues / portals in the sciences. These have been compiled into an excel spreadsheet (72 in total)
- The list has been priorized and 25 data archives / repositories / catalogues / portals in the sciences have been selected to be surveyed.
- Sherry Xie has begun this work.
- Barbara Craig and Tracey Laurialt will begin to develop a framework of analysis with the work to date at a full day meeting in Toronto Oct. 6, 2005

Study proposed by Barbara Craig, University of Toronto and Tracey Lauriault, Geomatics and Cartographic Research Centre at Vancouver BC Workshop in February of 2005 and approved by the Science Focus and the IP2 Director. Student resources have been allocated to this work.